On the forecasting horizon of seismicity models

David Rhoades
GNS Science, Lower Hutt, New Zealand

5th International Workshop on Statistical Seismology, Erice, 30 May – 6 June 2007

Introduction

- **Seismicity models often fitted and assessed based on instantaneous rate densities $\lambda(t,m,x,y)$, by maximising the log likelihood**

$$\ln L = \sum_{i=1}^{n} \ln \lambda(t_i, m_i, x_i, y_i) - E(N)$$

of the earthquake catalogue $\{(t_i,m_i,x_i,y_i),\ i=1,\ldots,n\}$ in a region.

- **Here, $\lambda(t,m,x,y)$ is typically evaluated using all prior information, however close to time t.**

Practical forecasting

In practice there is some future time-period of interest e.g.,

- next day
- next year
- next five years

Forecast must be made using only information available before the beginning of the period.

This may affect the performance of a model

Forecasting horizon

- **Let F_t denote information available up to but not including time t. Then define**

$$\lambda^0(t,m,x,y) = \lambda(t,m,x,y \mid F_t)$$

and

$$\lambda^\tau(t,m,x,y) = \lambda(t,m,x,y \mid F_{t-\tau})$$

- **At any time, λ^τ looks ahead to a forecasting horizon τ beyond the present.**

Information rate at a forecasting horizon

- A model's performance can be quantified by its *information rate*, $\Delta \ln L / n$, where

$$\Delta \ln L = \ln L - \ln L_0$$

where L and L_0 denote the likelihood under the model of interest and a reference model.

- Let the likelihood of the conditional intensity λ_τ with forecasting horizon τ be denoted $L(\tau)$. The information rate at forecasting horizon τ is then $\Delta \ln L(\tau) / n$, where

$$\Delta \ln L(\tau) = \ln L(\tau) - \ln L_0$$

Visibility at a forecasting horizon

$$\text{Visibility at forecasting horizon } \tau = \frac{\Delta \ln L(\tau)}{\Delta \ln L(0)}$$

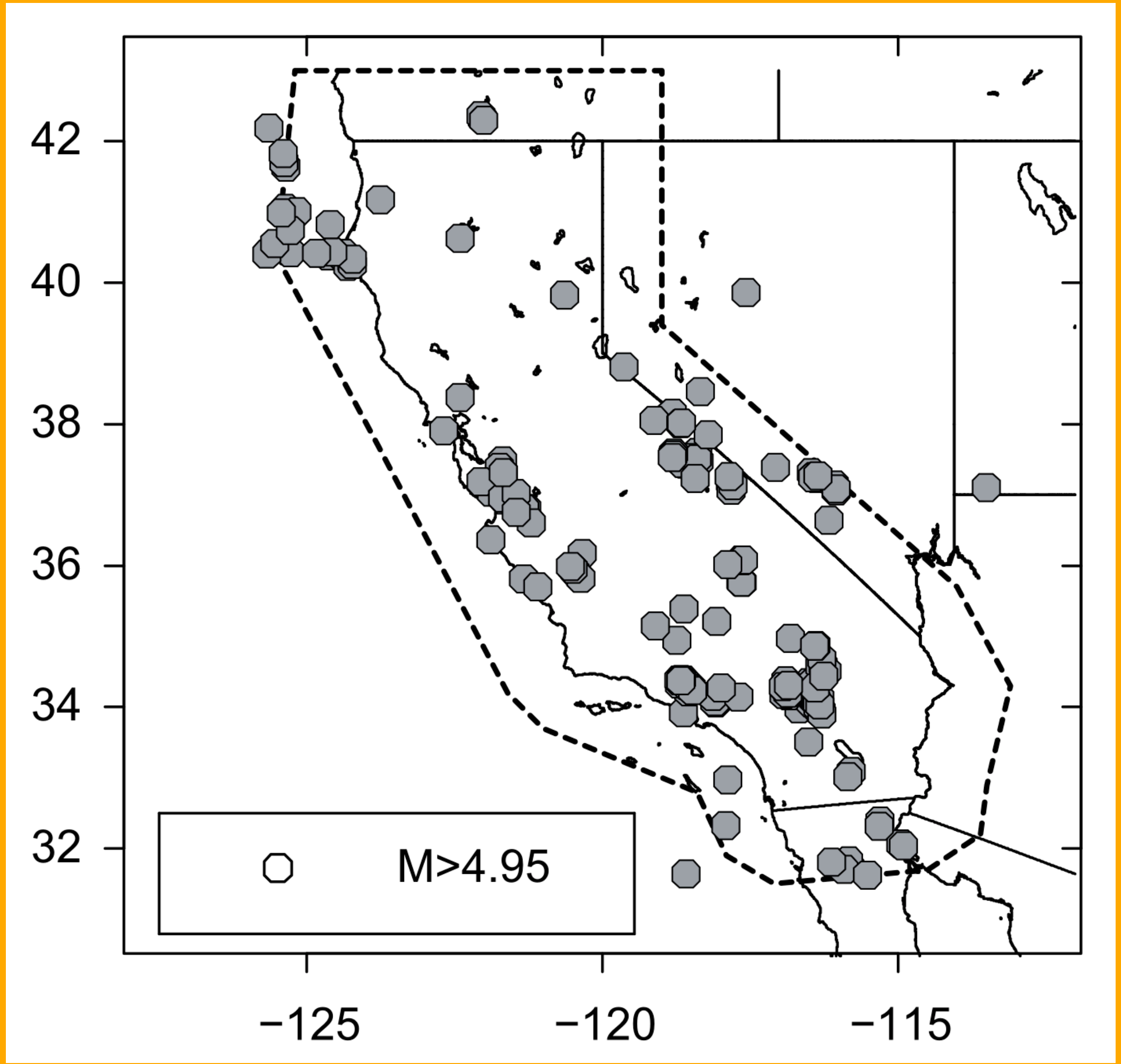
50% visibility horizon

$$\frac{\Delta \ln L(\tau_{50})}{\Delta \ln L(0)} = 0.5$$

Examples

- **SUP – stationary and spatially uniform Poisson model: static reference model.**
- **PPE – proximity to past earthquakes: quasi-static spatially varying reference model which updates as each new major earthquake occurs**
- **ETAS – time-space epidemic type aftershock model: short-term forecasting model which updates as each new earthquake occurs**
- **EEPAS – every earthquake a precursor according to scale: long-range forecasting model which updates as each new earthquake occurs**

Models were fitted to RELM test region, 1984-2004



m_c: Magnitude threshold for forecast

m_0: Magnitude threshold for data (EEPAS and ETAS)

Catalogue data from 1932 used.

PPE model

$$\lambda^{\tau}_{PPE}(t,m,x,y) = f_0(t)g_0(m)h_0(t,x,y)$$

where

$$f_0(t) = 1/(t - \tau - t_0),$$

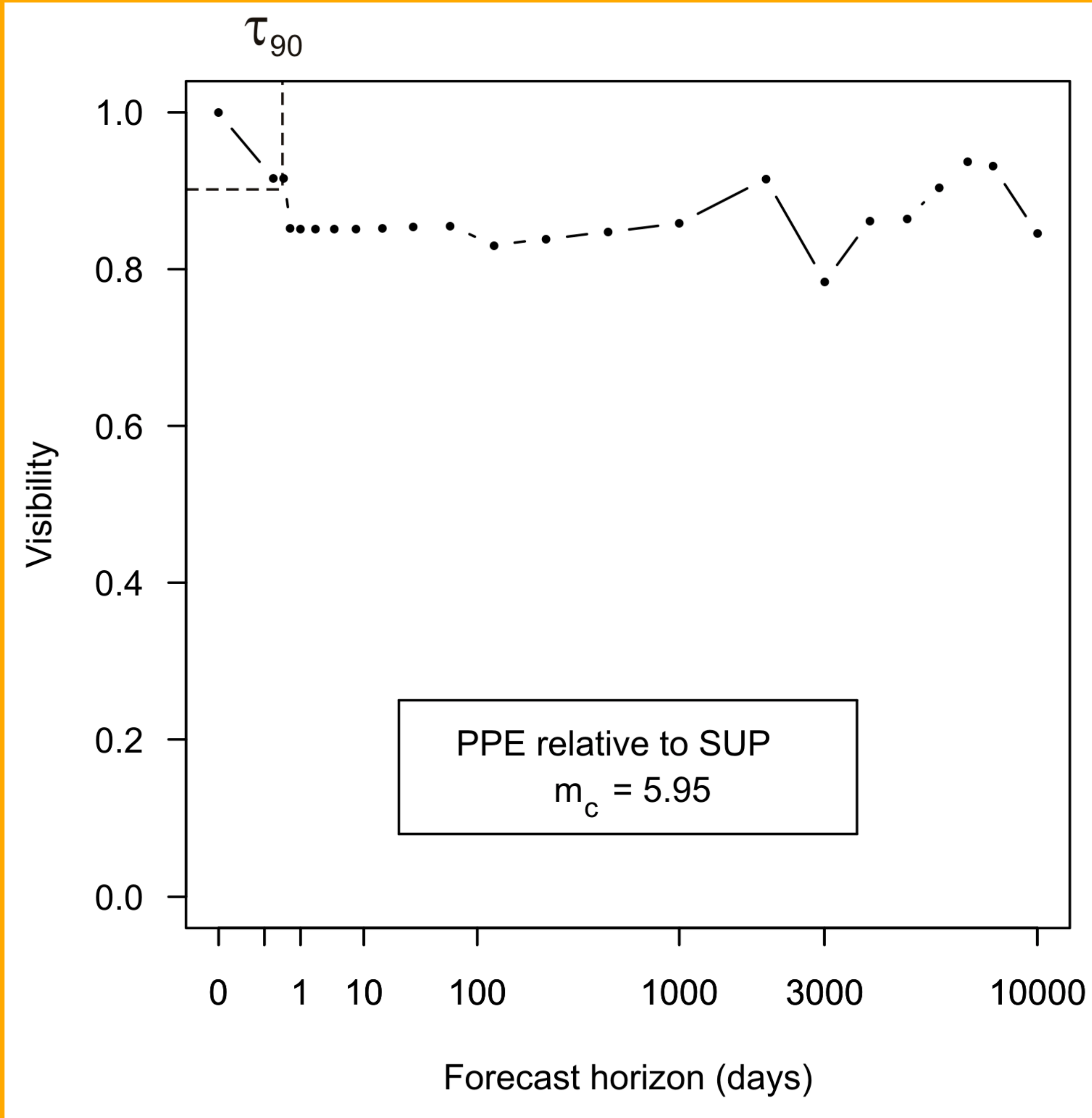
$$g_0(m) = \beta \exp\left[-\beta(m - m_c)\right] \quad (m > m_0),$$

$$h_0(t,x,y) = \sum_{t_i = t_0}^{t-\tau} h_{0i}(r_i), \quad \text{where}$$

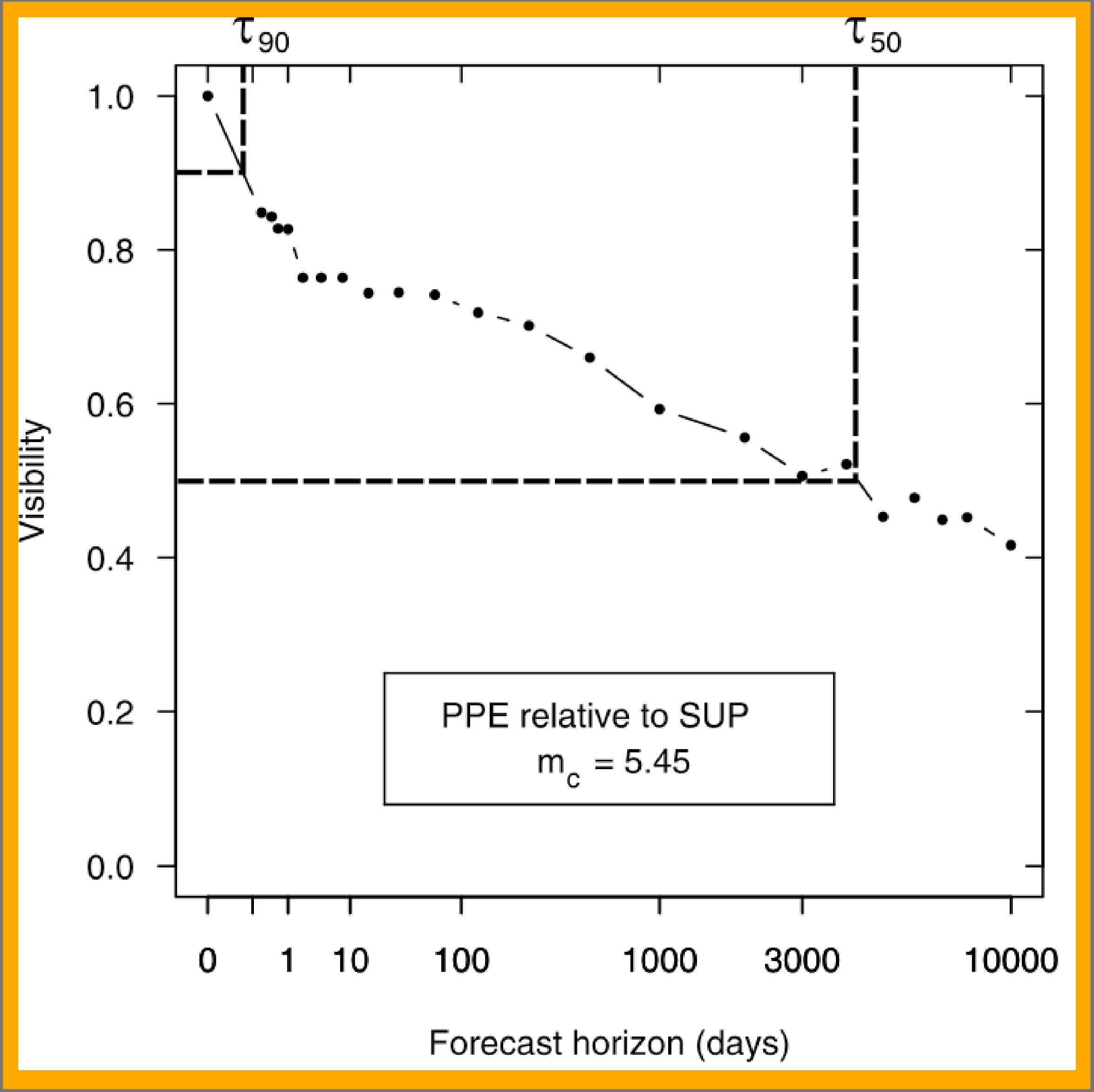
$$h_{0i}(r_i) = a(m_i - m_c)\frac{1}{\pi}\left(\frac{1}{d^2 + r_i^2}\right) + s$$

Refs : Jackson and Kagan (1999), Rhoades and Evison (2006)

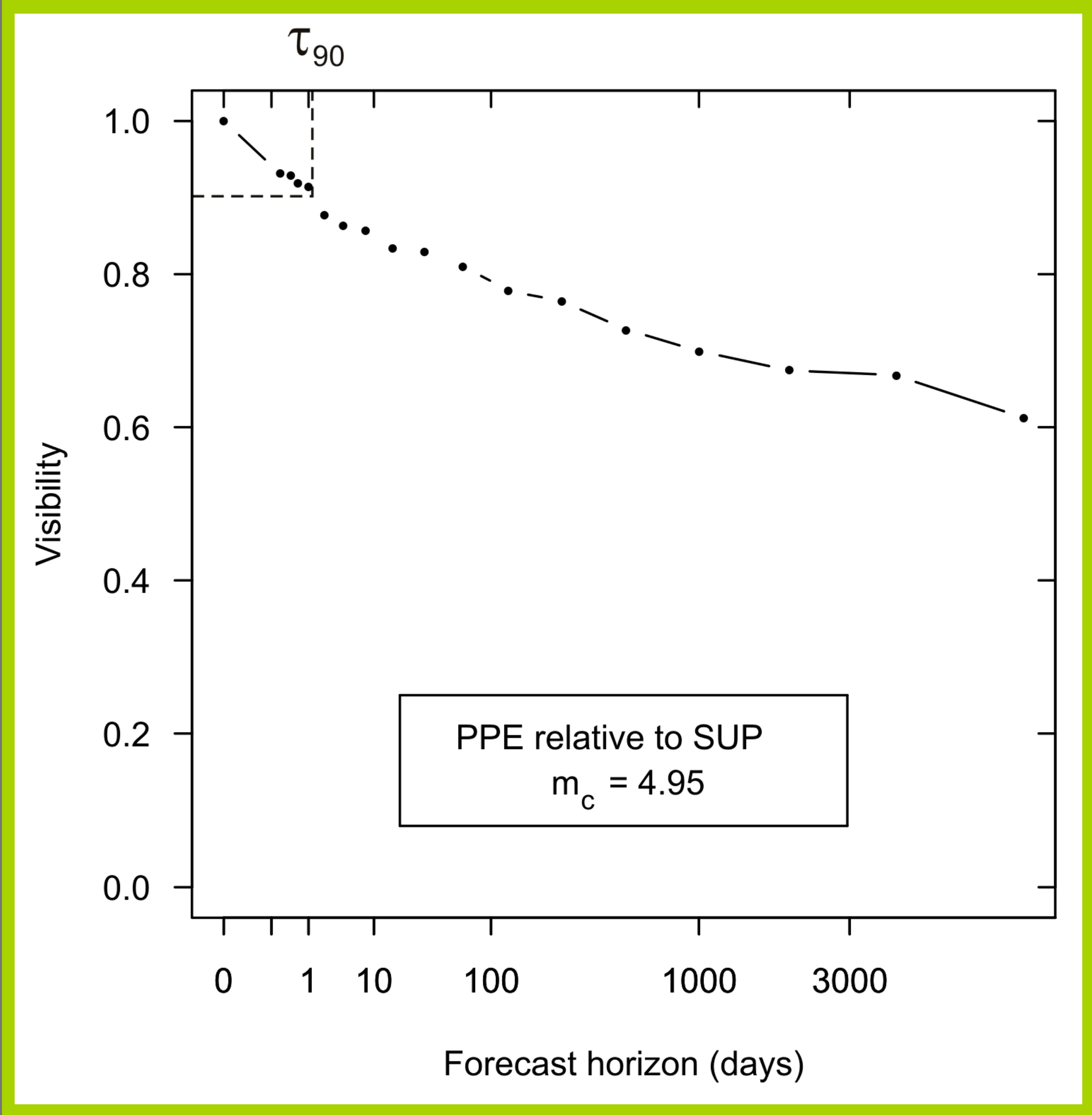
Visibility of PPE model relative to SUP



Visibility of PPE model relative to SUP



Visibility of PPE model relative to SUP



ETAS model

$$\lambda^{\tau}_{ETAS}(t,m,x,y) = \nu\lambda^{\tau}_{PPE}(t,m,x,y) + \sum_{t_i=t_0}^{t-\tau} f_{1i}(t)\,g_{1i}(m)\,h_{1i}(x,y)$$

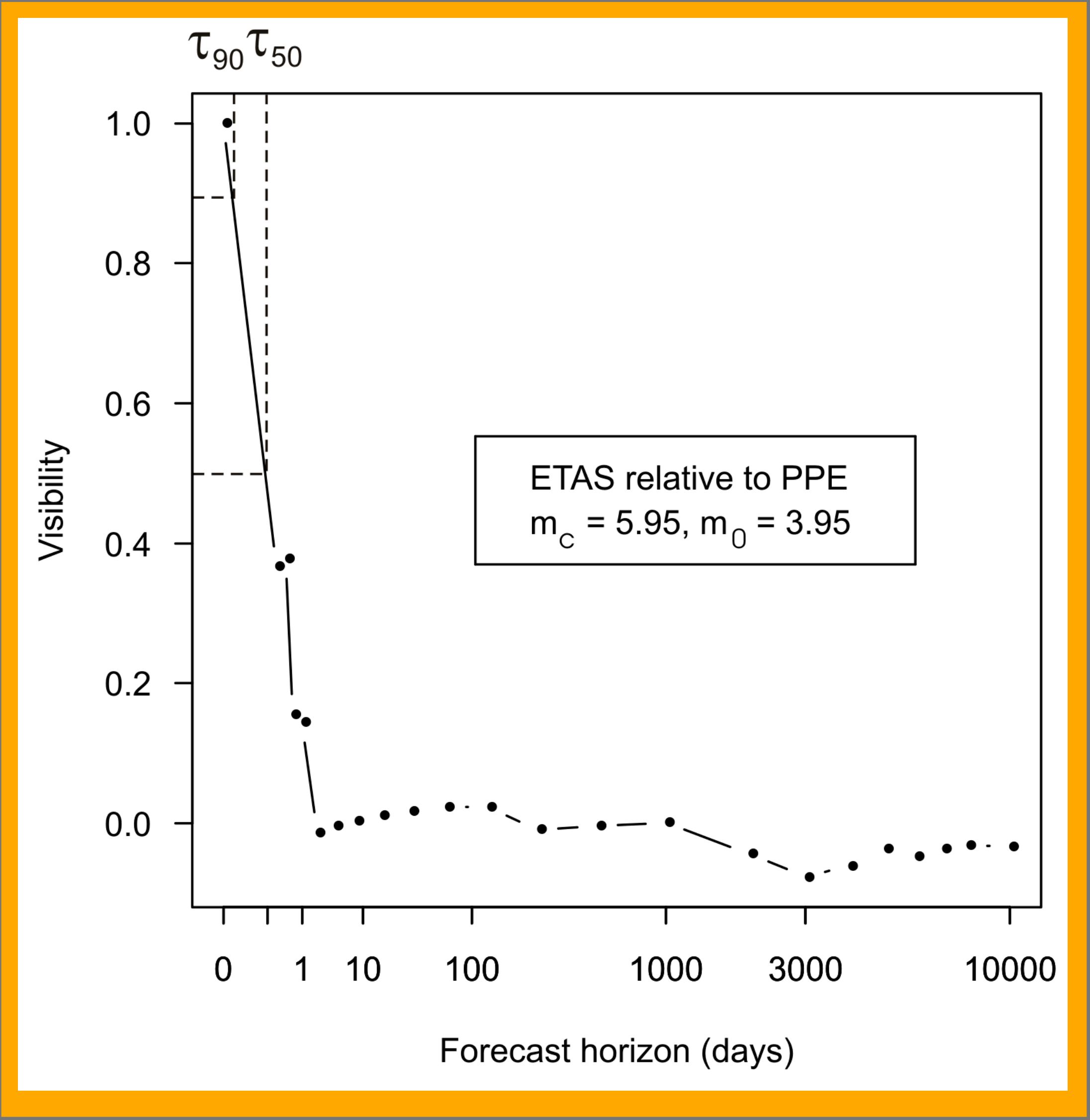
where

$$f_{1i}(t) = H(t-t_i)\frac{p-1}{(t-t_i+c)^p},$$

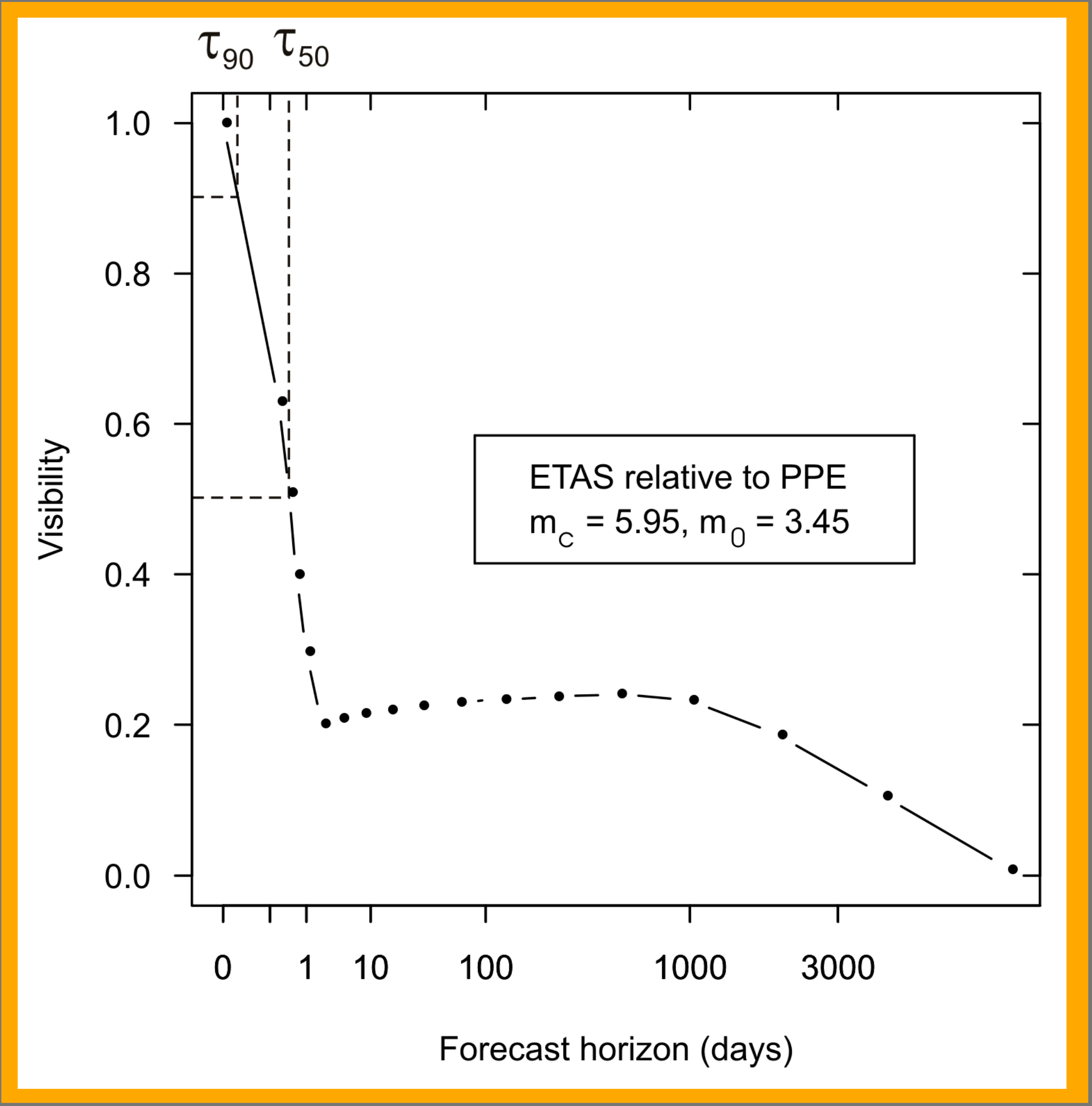
$$g_{1i}(m) = \beta\exp\left[-\beta(m-m_i)\right],$$

$$h_{1i}(x,y) = \frac{1}{2\pi\sigma^2 10^{m_i}}\exp\left[-\frac{(x-x_i)^2+(y-y_i)^2}{2\sigma^2 10^{m_i}}\right].$$

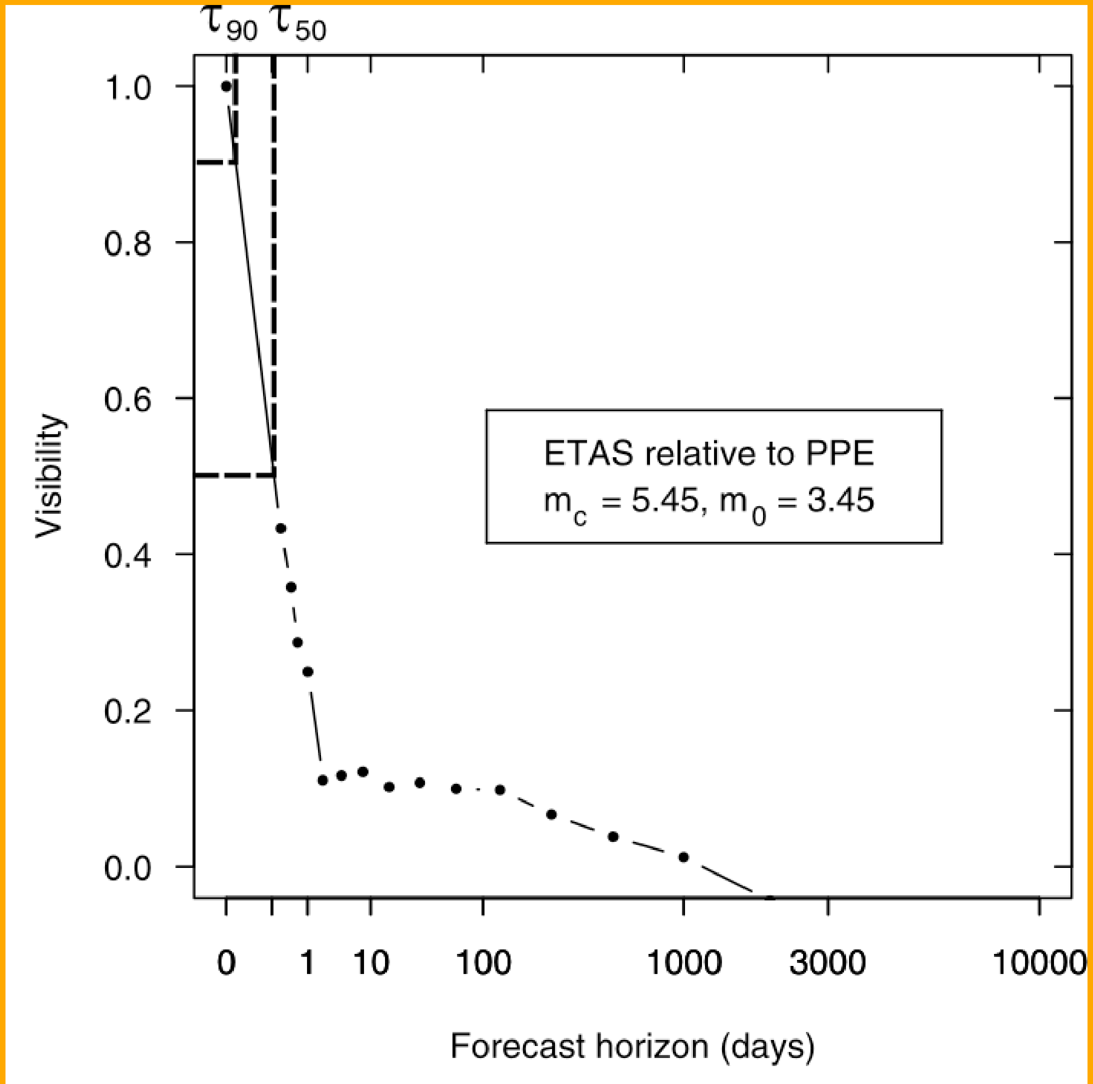
Visibility of ETAS relative to PPE at τ = 80 days



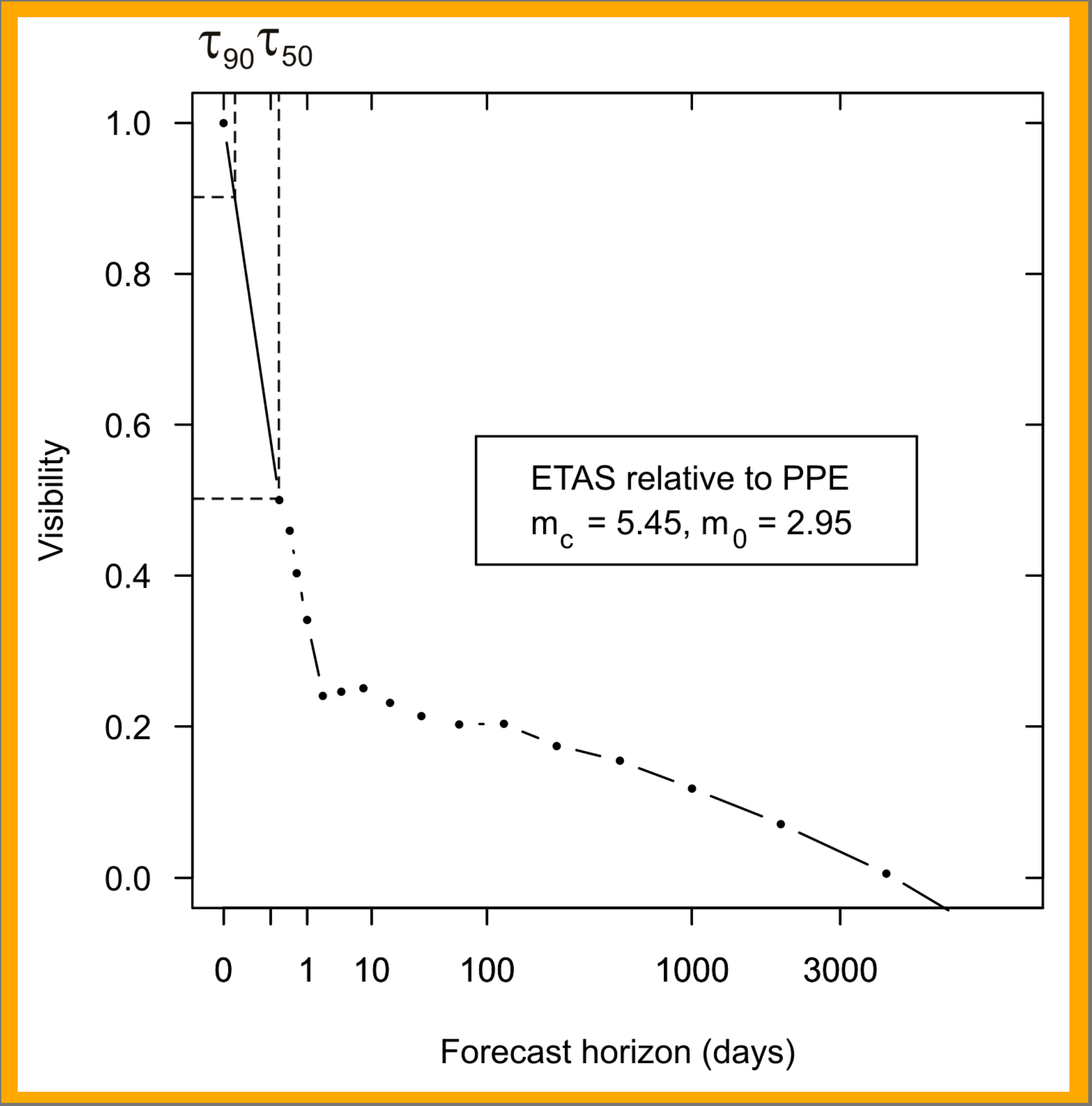
Visibility of ETAS relative to PPE at τ = 80 days



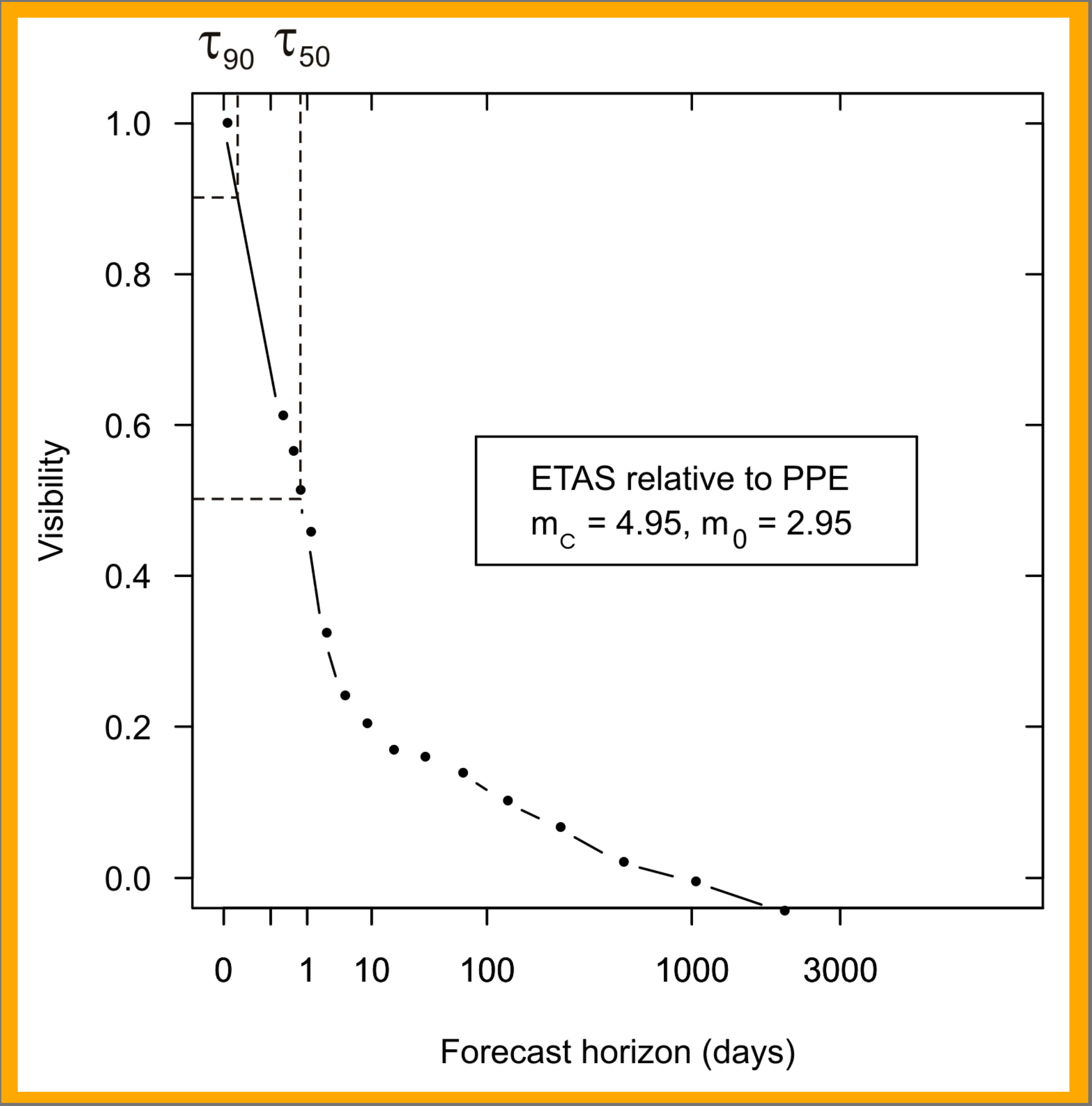
Visibility of ETAS relative to PPE at τ = 80 days



Visibility of ETAS relative to PPE at τ = 80 days



Visibility of ETAS relative to PPE at τ = 80 days



EEPAS model

$$\lambda^{\tau}_{EEPAS}(t,m,x,y) = \mu\lambda^{\tau}_{PPE}(t,m,x,y) + \sum_{t_i=t_0}^{t-\tau}\eta(m_i) f_{2i}(t) g_{2i}(m) h_{2i}(x,y)$$

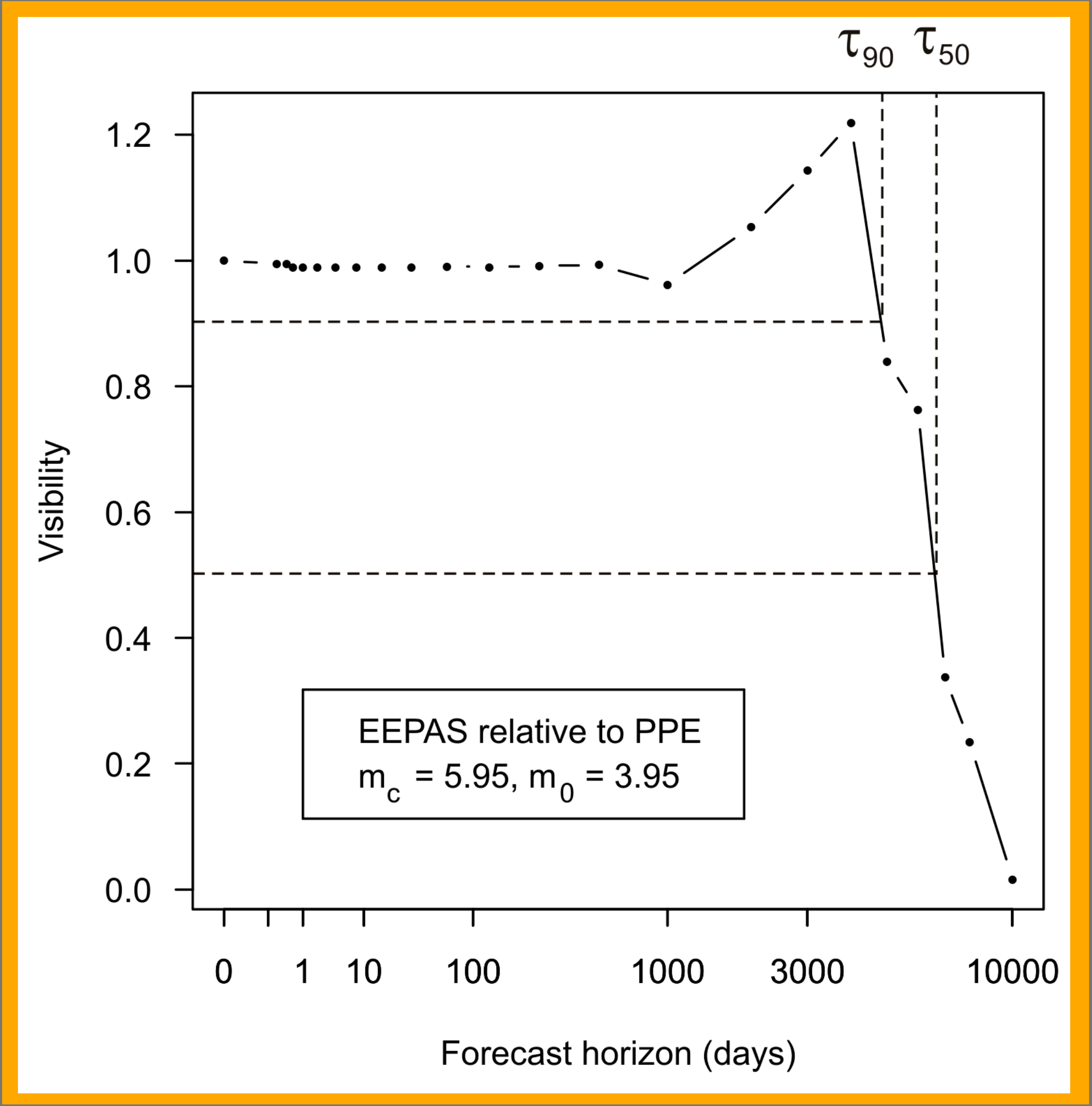
where $\eta(m_i)$ is a normalizat ion function, and

$$f_{2i}(t) = \frac{H(t-t_i)}{\sigma_M \ln(10)\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\log(t-t_i) - a_T - b_T m_i}{\sigma_T}\right)^2\right],$$

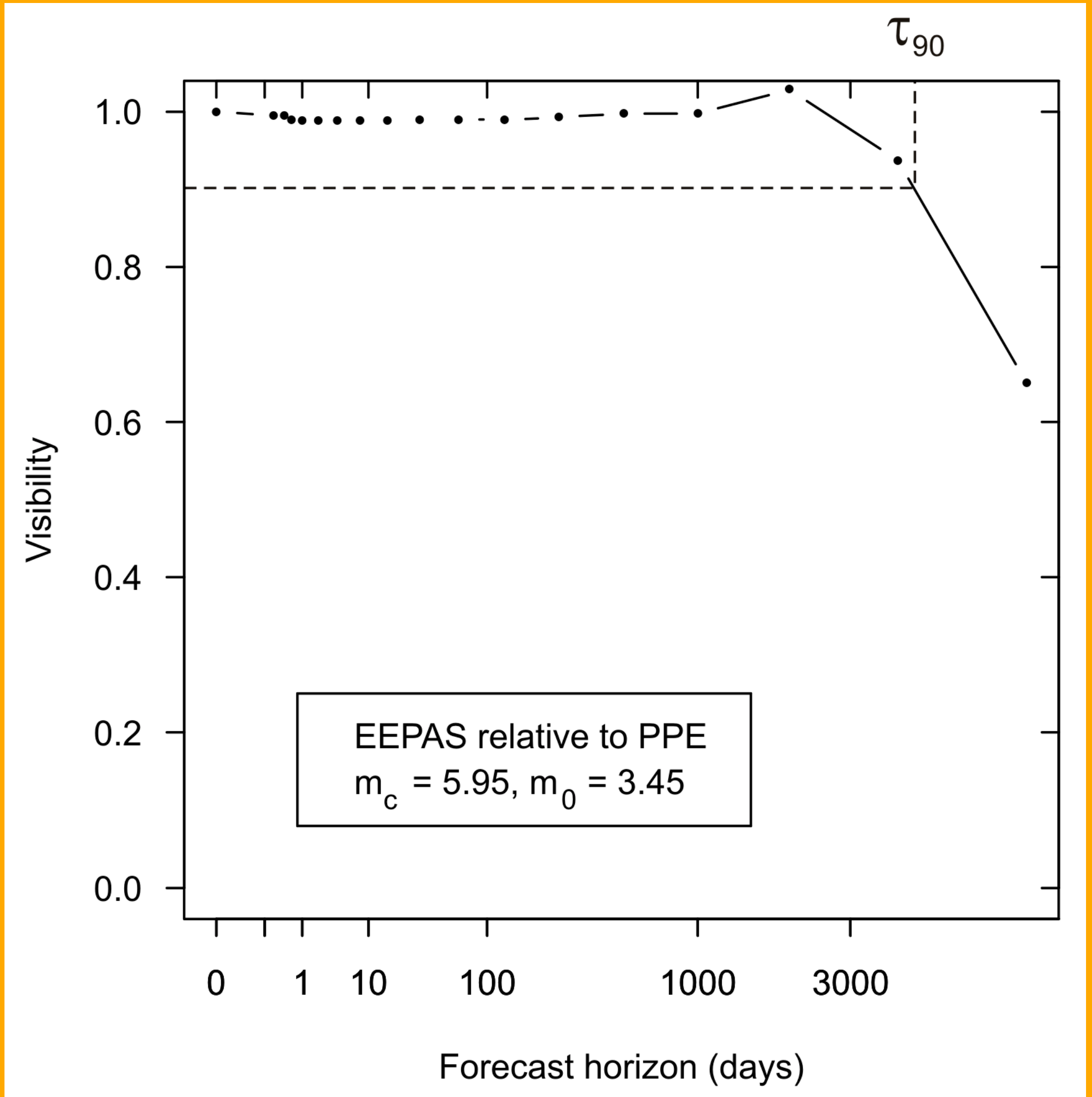
$$g_{2i}(m) = \frac{1}{\sigma_M\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{m - a_M - b_M m_i}{\sigma_m}\right)^2\right],$$

$$h_{2i}(x,y) = \frac{1}{2\pi\sigma_A^2 10^{b_A m_i}} \exp\left[-\frac{(x-x_i)^2 + (y-y_i)^2}{2\sigma_A^2 10^{b_A m_i}}\right].$$

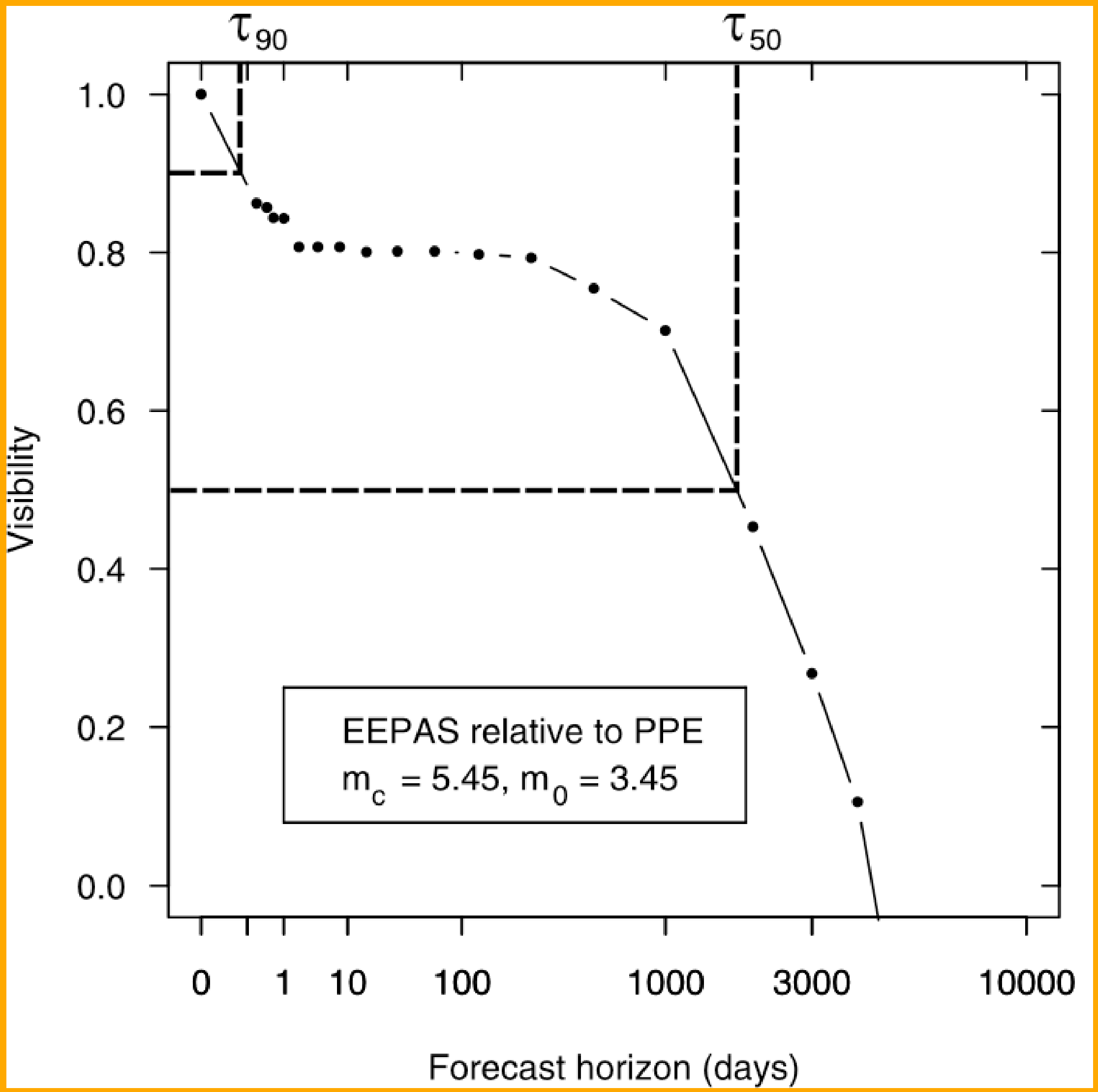
Visibility of EEPAS relative to PPE at τ = 80 days



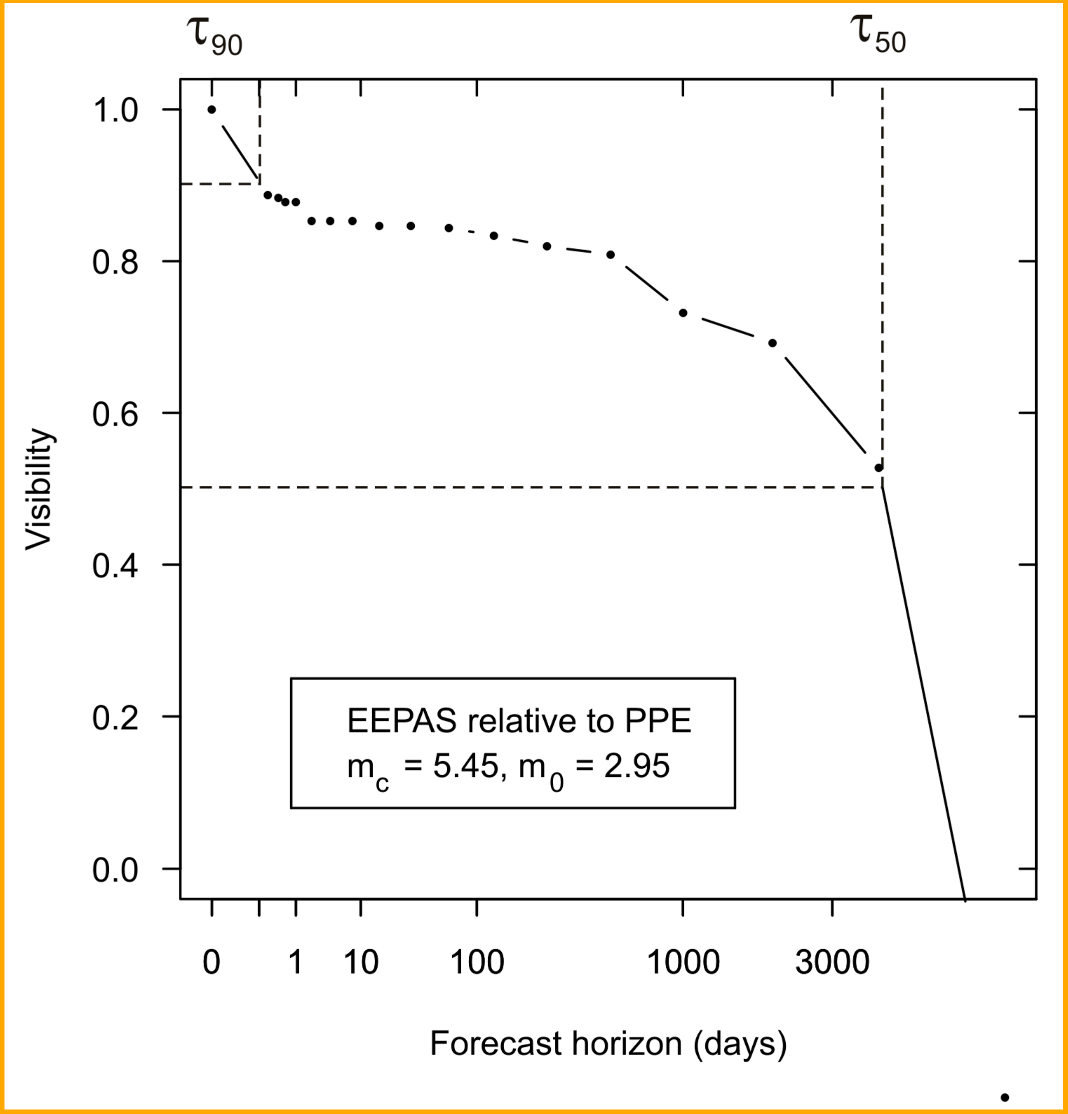
Visibility of EEPAS relative to PPE at τ = 80 days



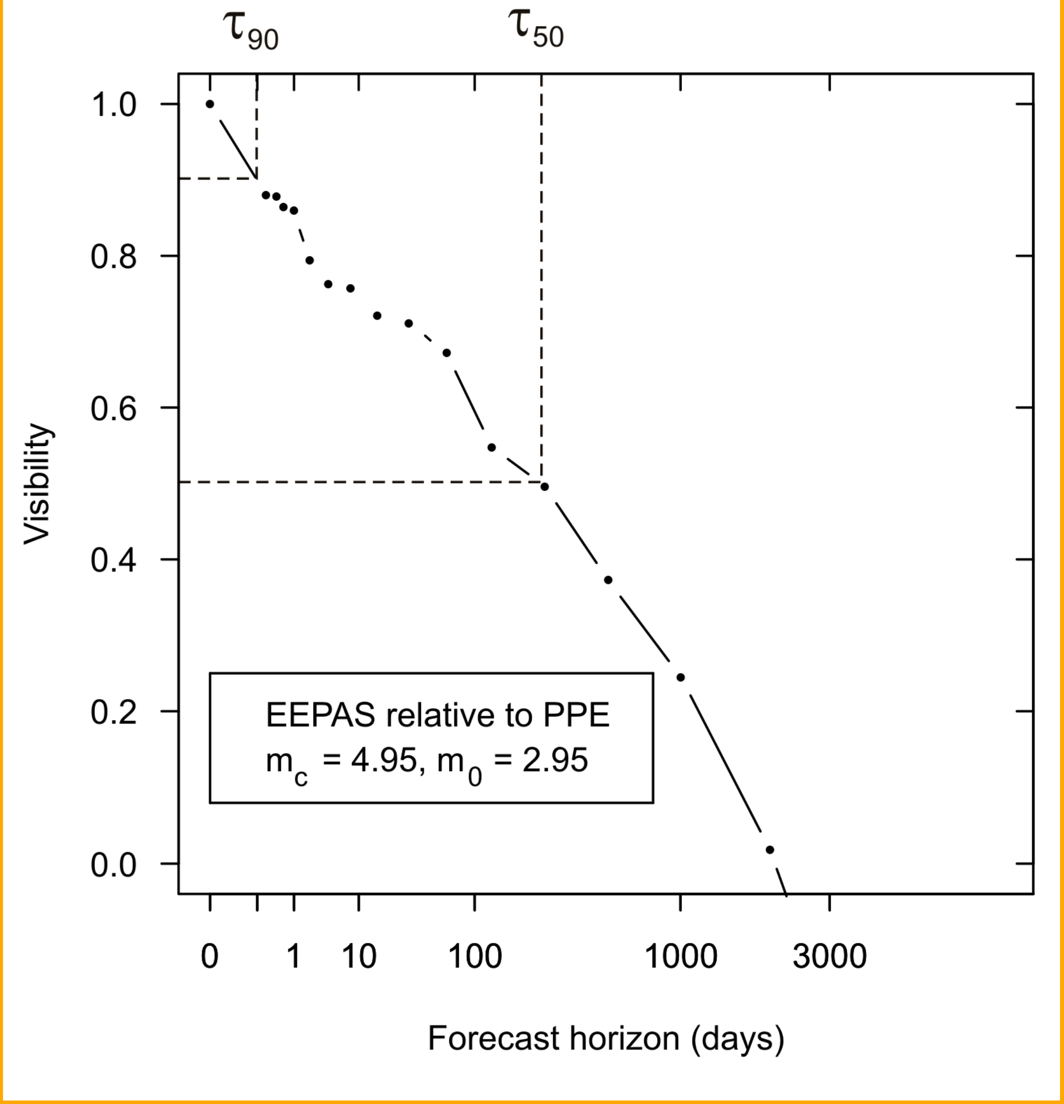
Visibility of EEPAS relative to PPE at τ = 80 days



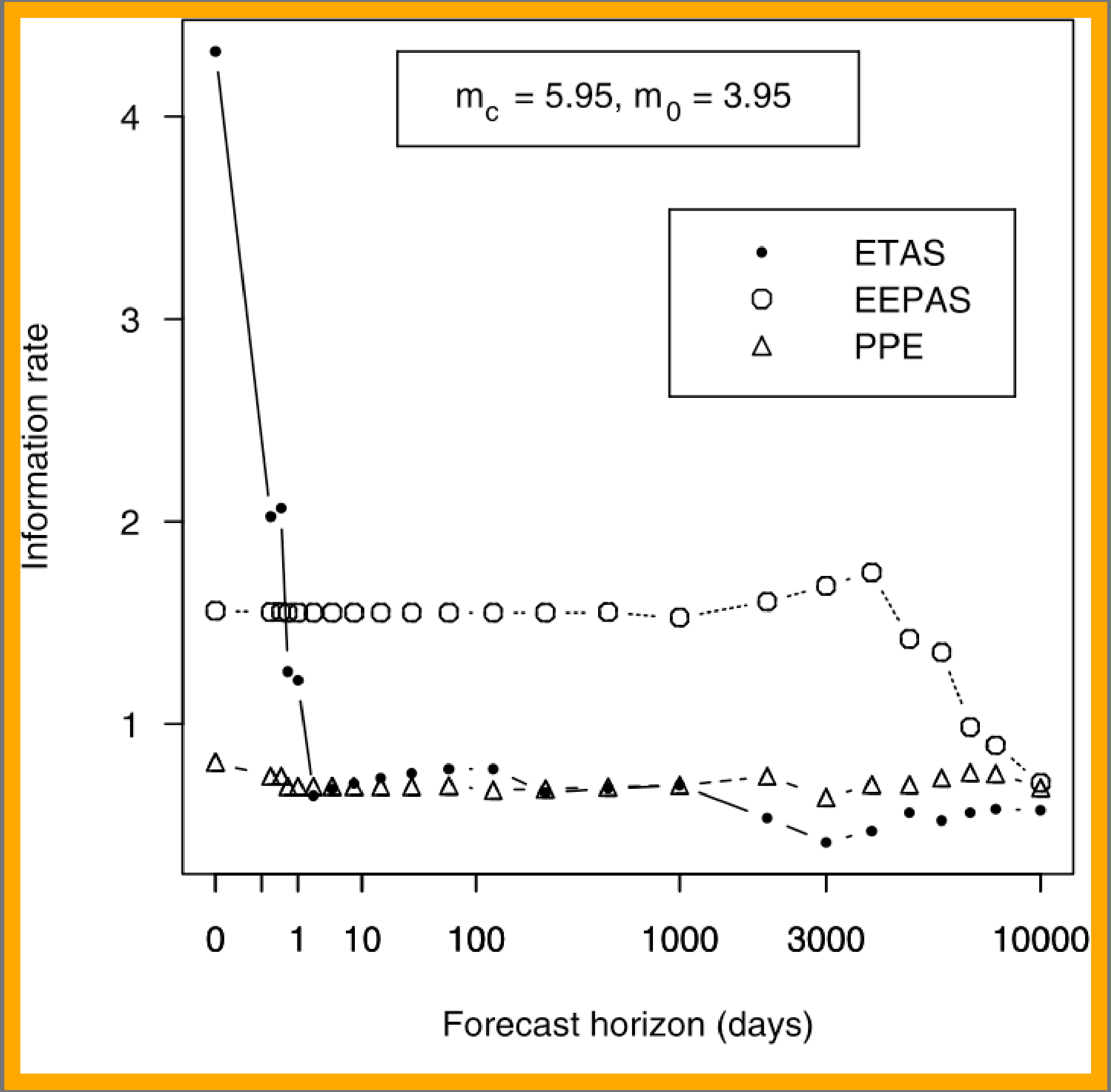
Visibility of EEPAS model relative to PPE at τ = 80 days



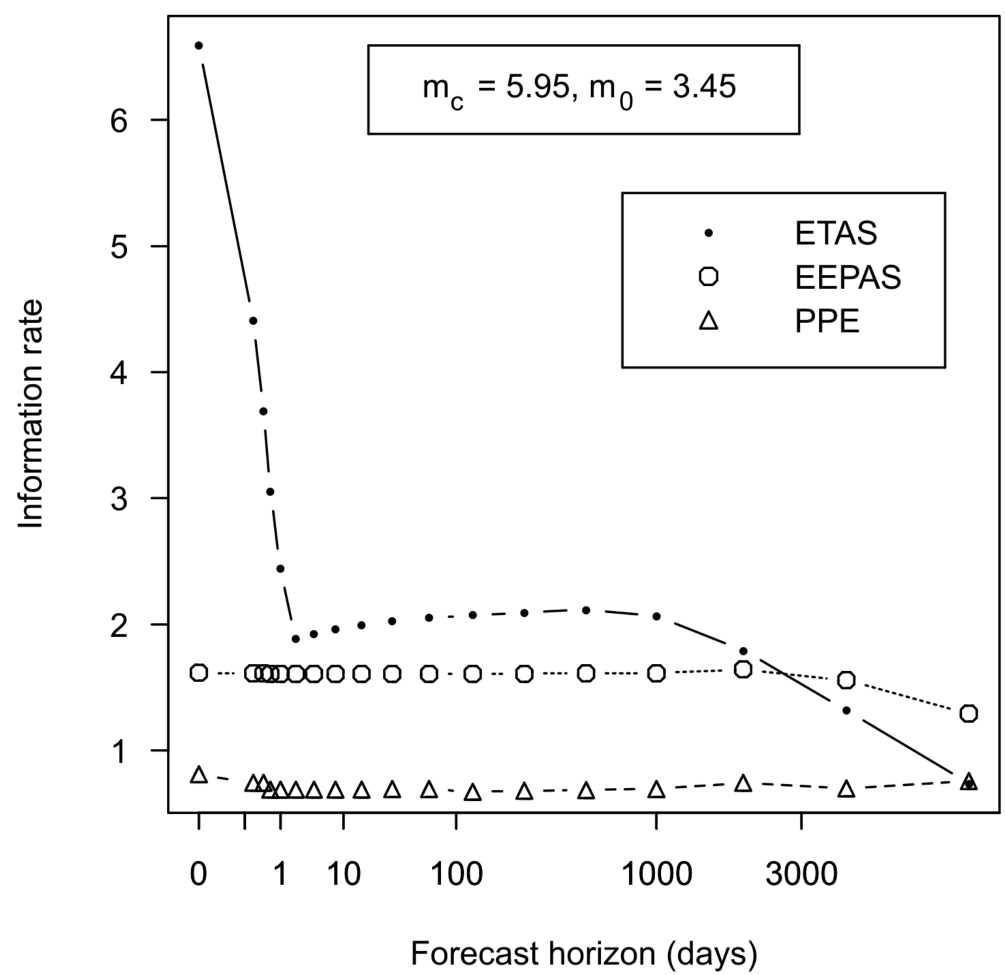
Visibility of EEPAS relative to PPE at τ = 80 days



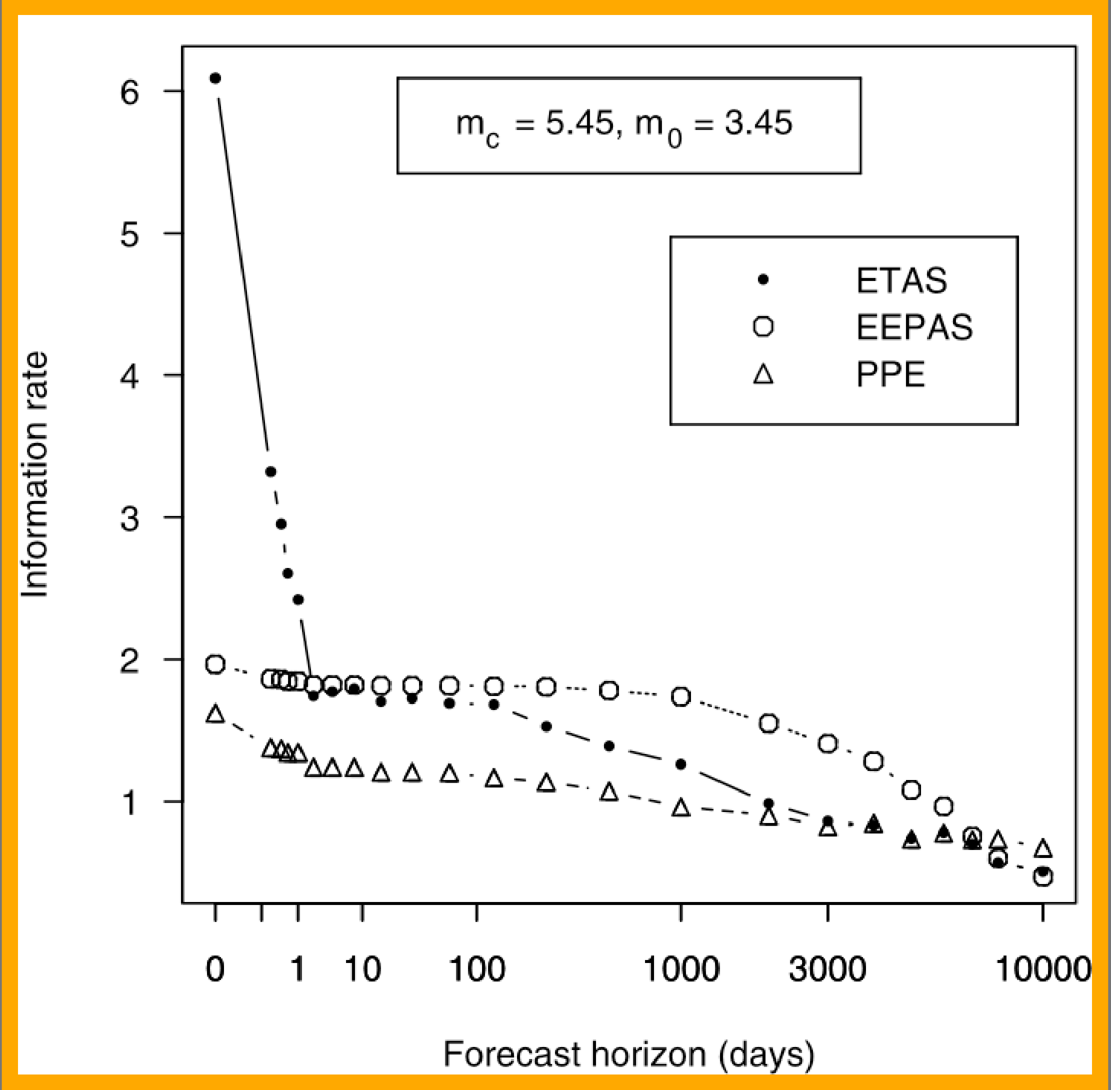
Comparison of information rates (relative to SUP)



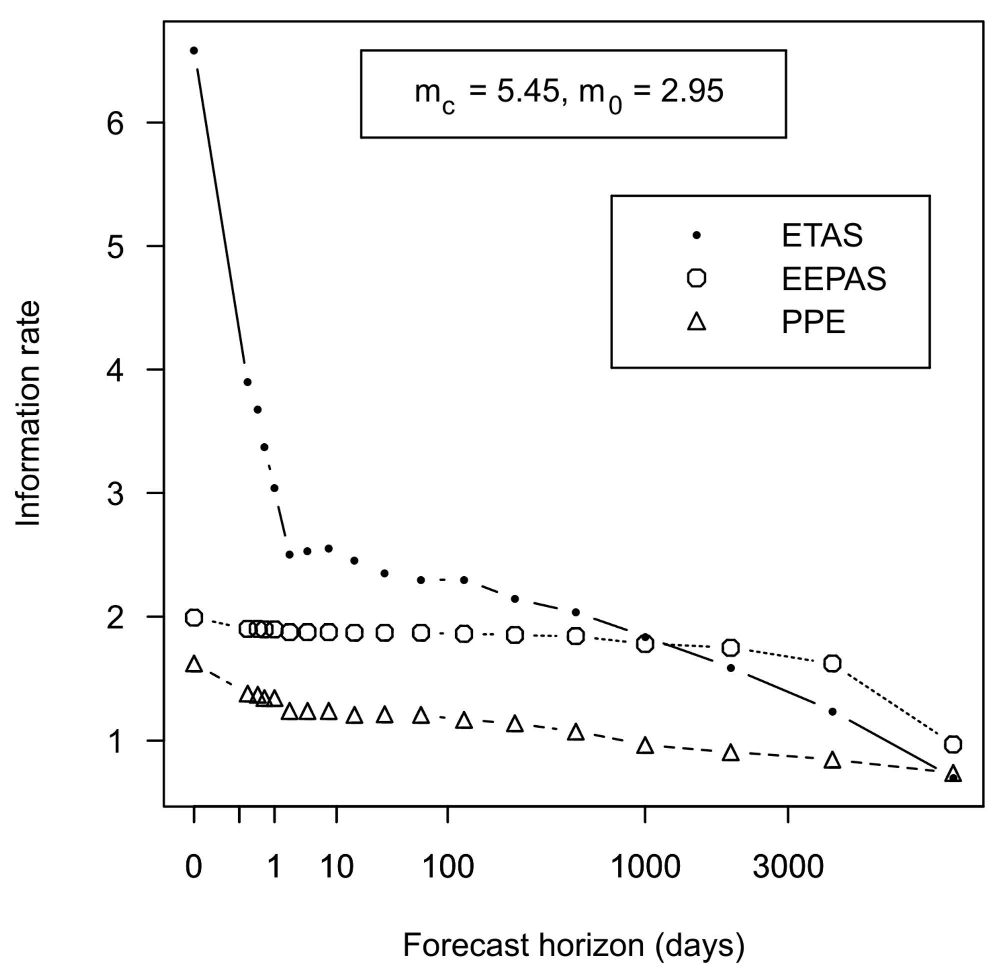
Comparison of information rates (relative to SUP)



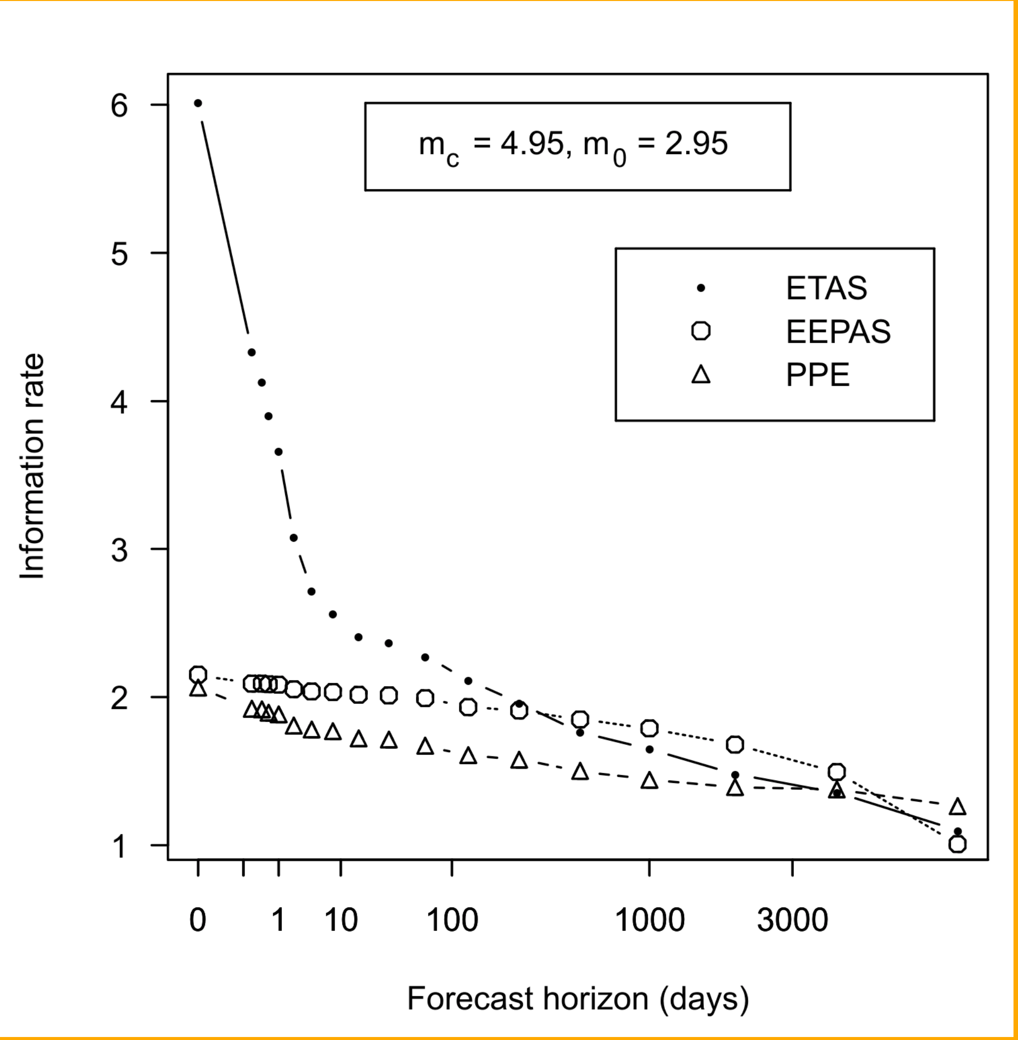
Comparison of information rates (relative to SUP)



Comparison of information rates (relative to SUP)



Comparison of information rates (relative to SUP)



Conclusion

- Analysis of the visibility of a model at different forecast horizons is a useful exploratory tool that can help us understand and improve existing seismicity models.
- It shows at what forecast horizons a model has predictive capability, and so is a guide to the usefulness of the model in practical forecasting.
- Comparison of information rates of different models as a function of forecast horizon may help us to construct optimal composite models for forecasting in a time-window of given length.